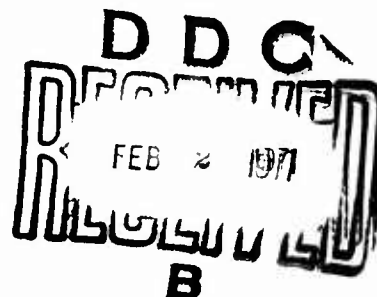AD717762

Research in Probability Theory

Final Report
Air Force Contract AF 49(638-1682)

Period Covering
February 1, 1966 to July 31, 1970

Roy L. Adler
Alan G. Konheim

IBM Thomas J. Watson Research Center
Yorktown Heights, New York

December 7, 1970

D D C
FEB 2 1971
B

This document has been approved for public
release and sale; its distribution is unlimited

Prepared

for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

UNITED STATES AIRFORCE

WASHINGTON, D. C.

13

INTRODUCTION

The following is a final report for Air Force Contract AF49 (638) -

1682 covering the period 1 February 1966 to 31 July 1970. The contract

supported research in probability theory and the personnel involved were

Dr. R. L. Adler and Dr. A. G. Konheim.

For the most part the activity supported dealt with two branches of

probability, one applied and the other pure. The applied one was concerned

with problems in the area of queueing and scheduling and the pure one with

ergodic theory. Dr. Konheim conducted the research in queueing and

Dr. Adler that in ergodic theory.

1.

A COMPREHENSIVE REVIEW

I. Ergodic Theory

A central problem in the study of dynamical systems is the question
of similarity, i.e., when are two systems equivalent under a change of
variables. This question which begins as a problem in analysis has been
shown as a result of this research to have its solution in many instances in
the subject of coding theory. For the continuous automorphisms of the two-
dimensional torus and the discrete time dynamical systems to which they
give rise the problem of equivalence under a change of variable was thorough-
ly investigated in two works, Entropy, a complete metric invariant for auto-
morphisms of the torus, and Similarity of automorphisms of the torus.
With such systems the key to finding whether a change of variables (which
has only to satisfy the weakest of requirements, measurability) exists lies
in the subject of information theory and a notion called entropy. It was
proven that two continuous ergodic automorphisms of the 2-torus are equiva-
lent under measurable changes of variable if and only if they have the same
entropy. The method of constructing the change of variable entailed rep-
resenting the history of orbits of points under the mapping as sequences
of symbols from a finite alphabet. Entropy is a concept that tells how
large such an alphabet should be and with what frequency the symbols
should occur. The symbolic sequences associated with torus automor-
phisms satisfy one-step transition rules, that is they come from finite

2.

state markov processes. A method of coding between these symbolic
sequences was developed that produces the required transformation
indicated in the theorem when entropy conditions are satisfied.

The significance of the above research to the Air Force is be-
lieved by this writer to be the following. Although direct applications
are removed, the subject of information theory has been advanced.
Some of the coding results may lead to contributions to the area of
cryptography and the security of data transmission systems. In another
direction the above work shows how information theory might possibly
be applied to some of the very hard problems of dynamical systems such
as the n-body problem and stability behavior of fluid flow.

II. Applied Probability

The research in this area was devoted to problems of queueing, scheduling, time sharing and sorting. In order to determine optimum policies a model has to be set up and its probabilistic nature analyzed. A sample of work in this area is the following.

In A Disk Access Model we consider a disk file and a buffer of capacity b.

4.
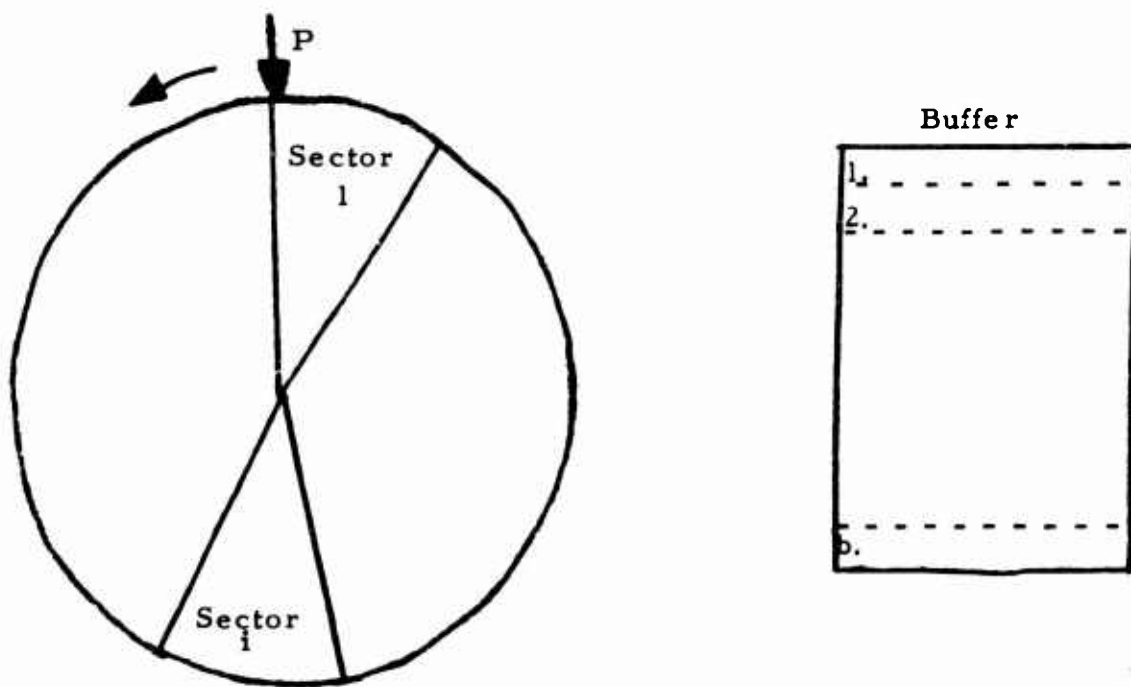
The disk is divided into m sectors. When a sector reaches the point P data can be read from the buffer into the sector. The disk access method is characterized by a parameter r; the number of requests to be read into a sector simultaneously. The parameter r takes values 1, 2, ..., b. The state of the buffer, at the start of a cycle (as shown above) is a vector

$$\underline{X} = (X^{(1)}, X^{(2)}, ..., X^{(m)})$$

where $X^{(i)}$ is the number of requests for sector 'i'. Here $X^{(i)} \geq 0$ $(1 \leq i \leq m)$ and $X^{(1)} + X^{(2)} + ... + X^{(m)} = b$. The set of all such vectors will be denoted by Ω_b. If there are t_i requests for sector i, when the disk is in the position to read into sector i, then min $(r, t_i) = s_i$ of them are satisfied. This removes s_i requests from the buffer and these are immediately replaced by s_i new requests (before the disk reaches the position in which it may read into the (i+1)st sector). They are chosen independently each with the uniform distribution over the set of integers $\{1, 2, ..., m\}$. The problem is to determine the stationary distribution of the state vector $\underline{X}$. In this paper we calculate the average rate in which data flows from the buffer to the disk for r = 1 and r = b.

The problem of sorting on a computer is one of the earliest problems in the data processing field. Sorting is the operation of arranging a sequence of records in some order. One might imagine the construction of a large telephone directory from several small ones.

It has been estimated that as much as 40% of all time on a computer in nonscientific applications is spent in sorting. In A Note on Merging the merging operation which occurs in sorting is studied. We imagine that a set of q 'strings' of numbers have been generated

$$X_{i, 1} \quad X_{i, 2} \cdots X_{i, n} \qquad (1 \le i \le q)$$

These 'strings' are in their natural order, i.e.

$$X_{i, 1} \le X_{i, 2} \le \cdots \le X_{i, n}$$

and we assume for simplicity that they are distinct. The q 'strings' are the result of internal sorting and n should be regarded as their average length. The q strings are to be merged to form a single list, this list being in the natural order. The particular merging operation is carried out on a disk system and the 'time' needed to merge is an idealization of the time needed in the disk system. In this note the expected length of time needed to merge is determined.

A question that arises in connection with the above work on sorting was dealt with in A Note on Order Statistics.

Let $X_1, X_2, \ldots$ be independent and identically distributed. The order statistics (of a sample of size n) is just the random variables $X_1, X_2, \ldots, X_n$ arranged in non-decreasing order

$$X_{1, n} \le X_{2, n} \le \cdots \le X_{n, n}$$

Here $X_{1, n} = \min(X_1, X_2, \ldots, X_n)$, $X_{n, n} = \max(X_1, X_2, \ldots, X_n)$. Let F be the common distribution function of the X_i. If F is the uniform

6.

distribution function   (on $[0, 1]$ say) then

$$(*) \qquad E(X_{i, c}) = \frac{1}{n + 1} \; .$$

It is natural to ask if (*) characterizes the uniform distribution on

$[0, 1]$. In this note we prove that the numbers $\{ E(X_{n, n}) \}_{n = 1}$ de-

termine F.

A Note on Growing Binary Search Trees is concerned with the

number of search operations need to locate an object in a collection.

A binary search tree is a tree (a graph without cycles) in which there

exists a distinguished vertex called <u>the root</u>. The root has degree two

(two edges leave the root) and all other vertices have degree one or

three (one edge enters and two leave). T will denote the family of

binary search trees and T_n the subset of T with n leaves. (A leaf

of a binary search tree is a vertex of degree one.) If $t \in T$ and ℓ is

a leaf then $d_t(\ell)$ is the distance of the leaf ℓ from the root. The

average leaf distance is given by

$$D_t = \frac{\displaystyle\sum_t d_t(\ell)}{\text{no. leaves in t}}$$

In this note we calculate $E(D_t: t \in T_n)$. The probability measure on T_n

is a 'natural' measure which one obtains by growing trees in T_n from

trees in T_{n-1}. The basic idea is to choose a leaf in a tree of T_{n-1}

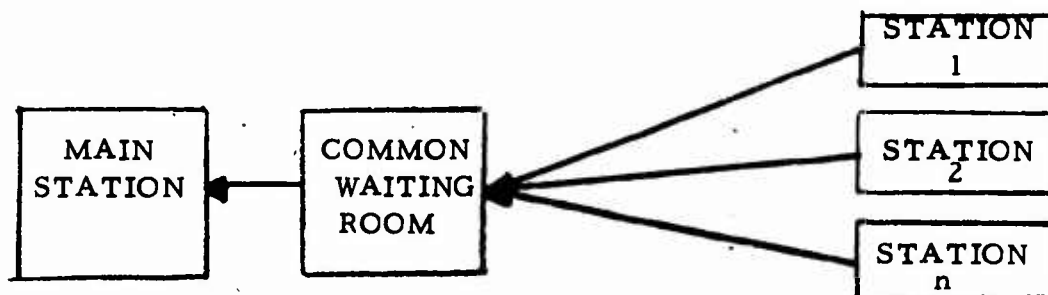and to change it into a vertex of degree three, thus obtaining an element

of T_n.

Finally Service in a Loop System studies the service problem
in two special data communications systems - the loop and star con-
figurations.
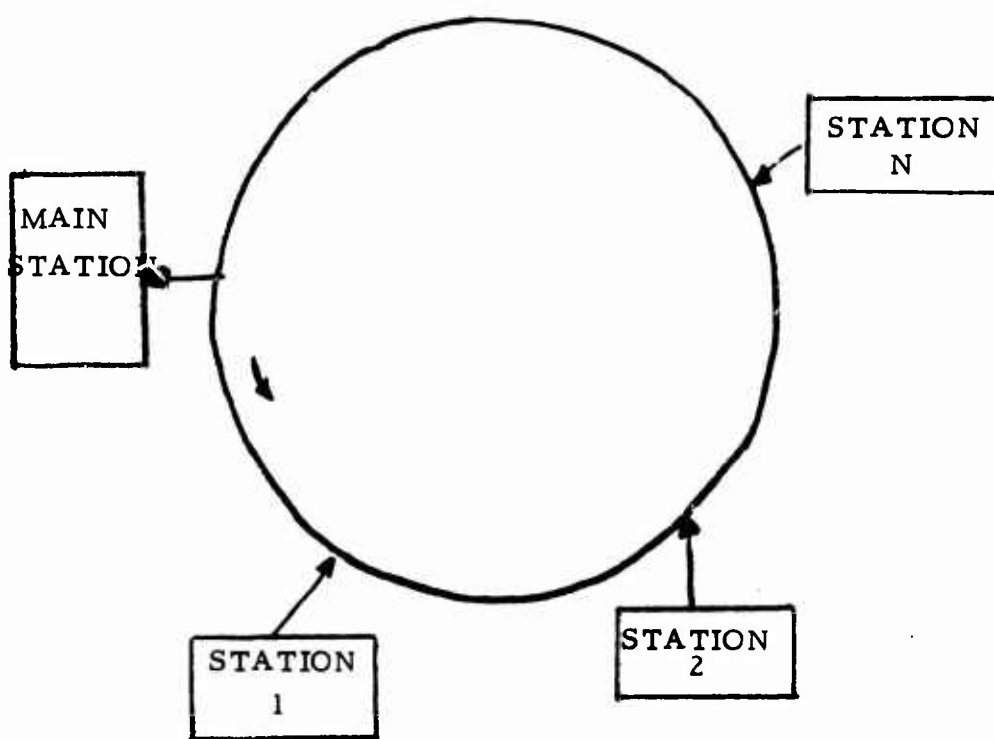
LOOP SYSTEM

STAR SYSTEM

8.

The Loop System consists of a main station (CPUX, N stations arranged on a loop (terminals) and a server of capacity C (a time multiplexed line with C frames per unit time period). The server makes a tour of the N stations serving requests which the stations make. The system is asynchronous in the sense that the full capacity of the server is available to the first station and thereafter the capacity offered to the succeeding stations is the residual capacity; the capacity C reduced by the number of requests served at preceeding stations. The input processes to the N stations are independent processes; each is a process with independent stationary increments.

The star system consists of the N stations each having a separate channel to a main waiting room (or buffer). The requests for service from the stations queue up in the buffer on a first-come-first served basis.

The problem analyzed in this paper is the waiting line and waiting time statistics for the loop system. By specializing the parameters the results for the star system are obtained. In particular the grade of service from station to station in the loop is of interest.

The value of this work to the Air Force coincides with its value to the computer industry. Any research which leads to more efficient methods of data processing will benefit both institutions.

9.

BIBLIOGRAPHY OF PUBLICATIONS AND
MANUSCRIPTS OF WORK PERFORMED

1. Adler, R. L. (with Weiss, B.), Entropy, "A Complete Metric Invariant for Automorphisms of the Torus", Proc. of Natl. Acad. of Sci. 57, 6, 1573-1576, (June, 1967).

2. Konheim, A. G. (with Chazan, D. and Weiss, B.), "A Note on Time Sharing", Journ. of Combinatorial Theory, 5, 344-369 (1968).

3. Konheim, A. G., "A Note on Time Sharing with Preferred Customers", Zeitschrift für Wahrscheinlichkeits Theorie and Verwandte Gebiete, 9, 112-130 (1968).

4. Konheim, A. G. (with Weiss, B.), "A Note On Functions Which Operate", Pacific Journ. of Math. (to appear).

5. Adler, R. L. (with Weiss, B.), "Similarity of Automorphisms of the Torus", Memoir #98, American Math. Soc. (1970).

6. Konheim, A. G., "A Note on Order Statistics", (submitted to Amer. Math. Monthly).

7. Konheim, A. G., "A Note on Merging", (submitted to SIAM Review).

8. Konheim, A. G., (with Newman, D. J.), "A Note on Growing Binary Search Trees", (submitted to the Journ. of Combinatorial Theory).

10.

9.   Konheim, A. G. (with Burge, W.),   "A Disk Access Model",
            (submitted to the Journ. of the Assoc. for computing
            machinery).

10.   Konheim, A. G. (Meister. B.),   "Service in a Loop System",
            (manuscript in preparation).

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
International Business Machines Corp. P. O. Box 218 Yorktown Heights, New York 10598	UNCLASSIFIED
	2b. GROUP

3 REPORT TITLE

RESEARCH IN PROBABILITY THEORY

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Scientific   Final

5 AUTHOR(S) *(First name, middle initial, last name)*

Roy L. Adler
Alan G. Konheim

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 7, 1970	10	10

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 49(638)-1682	
b. PROJECT NO 9749	
c. 61102F	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d. 681304	AFOSR 70-3017TR

10 DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11 SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research (NM) 1400 Wilson Boulevard Arlington, Virginia 22209

13 ABSTRACT

The activity supported dealt with two branches of probability, one applied and the other pure.

The applied one was concerned with problems in the area of queueing and scheduling.

The pure one was concerned with problems in ergodic theory.

DD FORM 1473
1 NOV 65